

1645

RECEIVED

NOV 23 2001

1600

TECH CENTER 1600/2900

## RAW SEQUENCE LISTING

DATE: 11/13/2001

PATENT APPLICATION: US/09/192,579B

TIME: 13:47:12

Input Set : A:\Sequence Listing.txt

Output Set: N:\CRF3\11132001\I192579B.raw

3 <110> APPLICANT: MENOZZI, Franco  
 4 LOCHT, Camille  
 6 <120> TITLE OF INVENTION: IDENTIFICATION AND CLONING OF A MYCOBACTERIAL ANTIGEN  
 7 CORRESPONDING TO A HEPARIN-BINDING HAEMAGGLUTININ  
 9 <130> FILE REFERENCE: 960-34  
 11 <140> CURRENT APPLICATION NUMBER: 09/192,579B  
 12 <141> CURRENT FILING DATE: 1998-11-17  
 14 <150> PRIOR APPLICATION NUMBER: FR 96 06168  
 15 <151> PRIOR FILING DATE: 1996-05-17  
 17 <160> NUMBER OF SEQ ID NOS: 20  
 19 <170> SOFTWARE: PatentIn Ver. 2.1  
 23 <210> SEQ ID NO: 1  
 24 <211> LENGTH: 39  
 25 <212> TYPE: PRT  
 26 <213> ORGANISM: Mycobacterium  
 28 <220> FEATURE:  
 29 <223> OTHER INFORMATION: sequence comprising a region involved in  
 30 interactions with sulphated glycoconjugates and in  
 31 heparin binding  
 33 <400> SEQUENCE: 1  
 34 Lys Lys Ala Ala Pro Ala Lys Lys Ala Ala Pro Ala Lys Lys Ala Ala  
 35 1 5 10 15  
 37 Pro Ala Lys Lys Ala Ala Lys Lys Ala Pro Ala Lys Lys Ala Ala  
 38 20 25 30  
 40 Ala Lys Lys Val Thr Gln Lys  
 41 35  
 44 <210> SEQ ID NO: 2  
 45 <211> LENGTH: 10  
 46 <212> TYPE: PRT  
 47 <213> ORGANISM: Mycobacterium  
 49 <220> FEATURE:  
 50 <223> OTHER INFORMATION: peptide S1441  
 52 <400> SEQUENCE: 2  
 53 Lys Ala Glu Gly Tyr Leu Glu Ala Ala Thr  
 54 1 5 10  
 58 <210> SEQ ID NO: 3  
 59 <211> LENGTH: 18  
 60 <212> TYPE: PRT  
 61 <213> ORGANISM: Mycobacterium  
 63 <220> FEATURE:  
 64 <221> NAME/KEY: CDS  
 65 <222> LOCATION: (1)  
 66 <223> OTHER INFORMATION: peptide S1443; Xaa can be any amino acid  
 68 <400> SEQUENCE: 3  
 W--> 69 Xaa Glu Gly Tyr Val Asp Gln Ala Val Glu Leu Thr Gln Glu Ala Leu  
 70 1 5 10 15  
 72 Gly Lys

ENTERED

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/192,579B

DATE: 11/13/2001

TIME: 13:47:12

Input Set : A:\Sequence Listing.txt

Output Set: N:\CRF3\11132001\I192579B.raw

```

76 <210> SEQ ID NO: 4
77 <211> LENGTH: 9
78 <212> TYPE: PRT
79 <213> ORGANISM: Mycobacterium
81 <220> FEATURE:
82 <221> NAME/KEY: CDS
83 <222> LOCATION: (1), (4) and (8)
84 <223> OTHER INFORMATION: peptide S1446; Xaa can be any amino acid
86 <400> SEQUENCE: 4
W--> 87 Xaa Gln Glu Xaa Leu Pro Glu Xaa Leu
88      1          5
91 <210> SEQ ID NO: 5
92 <211> LENGTH: 7
93 <212> TYPE: PRT
94 <213> ORGANISM: Mycobacterium
96 <220> FEATURE:
97 <223> OTHER INFORMATION: Peptide S1447
99 <400> SEQUENCE: 5
100 Phe Thr Ala Glu Glu Leu Arg
101      1          5
104 <210> SEQ ID NO: 6
105 <211> LENGTH: 17
106 <212> TYPE: DNA
107 <213> ORGANISM: Mycobacterium
109 <220> FEATURE:
110 <223> OTHER INFORMATION: Oligonucleotide originated from the S1441 peptide
111      (oligo S1441)
113 <400> SEQUENCE: 6
114 aaggcsgagg gstacct                               17
117 <210> SEQ ID NO: 7
118 <211> LENGTH: 17
119 <212> TYPE: DNA
120 <213> ORGANISM: Mycobacterium
122 <220> FEATURE:
123 <223> OTHER INFORMATION: Oligonucleotide originated from the S1441 peptide
124      (reverse oligo S1441)
126 <400> SEQUENCE: 7
127 aggtascctt csgcctt                               17
130 <210> SEQ ID NO: 8
131 <211> LENGTH: 17
132 <212> TYPE: DNA
133 <213> ORGANISM: Mycobacterium
135 <220> FEATURE:
136 <223> OTHER INFORMATION: Oligonucleotide originated from the S1443 peptide
137      (oligo S1443)
139 <400> SEQUENCE: 8
140 gaccaggcsg tsgagct                               17
143 <210> SEQ ID NO: 9
144 <211> LENGTH: 17

```

## RAW SEQUENCE LISTING

DATE: 11/13/2001

PATENT APPLICATION: US/09/192,579B

TIME: 13:47:12

Input Set : A:\Sequence Listing.txt

Output Set: N:\CRF3\11132001\I192579B.raw

```

145 <212> TYPE: DNA
146 <213> ORGANISM: Mycobacterium
148 <220> FEATURE:
149 <223> OTHER INFORMATION: Oligonucleotide originated from the S1443 peptide
150     (reverse oligo S1443)
152 <400> SEQUENCE: 9
153 agctcsacs g cctggtc 17
156 <210> SEQ ID NO: 10
157 <211> LENGTH: 21
158 <212> TYPE: DNA
159 <213> ORGANISM: Mycobacterium
161 <220> FEATURE:
162 <223> OTHER INFORMATION: Oligonucleotide named HBHASEq1 and used for
163     sequencing the gene coding for HBHA
165 <400> SEQUENCE: 10
166 agccggtaca acgagctggt c 21
169 <210> SEQ ID NO: 11
170 <211> LENGTH: 21
171 <212> TYPE: DNA
172 <213> ORGANISM: Mycobacterium
174 <220> FEATURE:
175 <223> OTHER INFORMATION: Oligonucleotide named HBHA Seq1inv and used for
176     sequencing the gene coding for HBHA
178 <400> SEQUENCE: 11
179 gaccagctcg ttgtaccggc t 21
182 <210> SEQ ID NO: 12
183 <211> LENGTH: 19
184 <212> TYPE: DNA
185 <213> ORGANISM: Mycobacterium
187 <220> FEATURE:
188 <223> OTHER INFORMATION: Oligonucleotide named HBHASEq2 and used for
189     sequencing the gene coding for HBHA
191 <400> SEQUENCE: 12
192 catccaacac gtcgactcc 19
195 <210> SEQ ID NO: 13
196 <211> LENGTH: 19
197 <212> TYPE: DNA
198 <213> ORGANISM: Mycobacterium
200 <220> FEATURE:
201 <223> OTHER INFORMATION: Oligonucleotide named HBHA Seq3 and used for
202     sequencing the gene coding for HBHA
204 <400> SEQUENCE: 13
205 ttgatgtcat caatgttcg 19
208 <210> SEQ ID NO: 14
209 <211> LENGTH: 19
210 <212> TYPE: DNA
211 <213> ORGANISM: Mycobacterium
213 <220> FEATURE:
214 <223> OTHER INFORMATION: Oligonucleotide named HBHA Seq4 and used for

```

## RAW SEQUENCE LISTING

DATE: 11/13/2001

PATENT APPLICATION: US/09/192,579B

TIME: 13:47:12

Input Set : A:\Sequence Listing.txt

Output Set: N:\CRF3\11132001\I192579B.raw

```

215      sequencing the gene coding for HBHA
217 <400> SEQUENCE: 14
218 cgtggaccag gcggtggag                      19
221 <210> SEQ ID NO: 15
222 <211> LENGTH: 21
223 <212> TYPE: DNA
224 <213> ORGANISM: Mycobacterium
226 <220> FEATURE:
227 <223> OTHER INFORMATION: Oligonucleotide named HBHA Seq 5 and used for
228      sequencing the gene coding for HBHA
230 <400> SEQUENCE: 15
231 gacgatcagg aggtttcccc g                      21
234 <210> SEQ ID NO: 16
235 <211> LENGTH: 24
236 <212> TYPE: DNA
237 <213> ORGANISM: Mycobacterium
239 <220> FEATURE:
240 <223> OTHER INFORMATION: Oligonucleotide named reverse primer and used for
241      sequencing the gene coding for HBHA
243 <400> SEQUENCE: 16
244 agcggataac aatttcacac agga                      24
247 <210> SEQ ID NO: 17
248 <211> LENGTH: 149
249 <212> TYPE: DNA
250 <213> ORGANISM: Mycobacterium
252 <220> FEATURE:
253 <223> OTHER INFORMATION: nucleotide sequence and amino sequence of a fragment of HBHA
254      deduced from a PCR fragment of chromosomal BCG DNA
256 <220> FEATURE:
257 <221> NAME/KEY: CDS
258 <222> LOCATION: (1)..(147)
260 <400> SEQUENCE: 17
261 aag gcc gag ggc tac ctc gag gcc gcg act agc cgg tac aac gag ctg      48
262 Lys Ala Glu Gly Tyr Leu Glu Ala Ala Thr Ser Arg Tyr Asn Glu Leu
263 1      5      10      15
265 gtc gag cgc ggt gag gcc gct cta gag cgg ctg cgc agc cag cag agc      96
266 Val Glu Arg Gly Glu Ala Ala Leu Glu Arg Leu Arg Ser Gln Gln Ser
267      20      25      30
269 ttc gag gaa gtg tcg gcg ccc gcc gaa ggc tac gtg gac cag gcg gtc      144
270 Phe Glu Glu Val Ser Ala Pro Ala Glu Gly Tyr Val Asp Gln Ala Val
271      35      40      45
273 gag ct                      149
274 Glu
277 <210> SEQ ID NO: 18
278 <211> LENGTH: 49
279 <212> TYPE: PRT
280 <213> ORGANISM: Mycobacterium
281 <223> OTHER INFORMATION: amino sequence of a fragment of HBHA deduced from a
286 <400> SEQUENCE: 18

```

## RAW SEQUENCE LISTING

DATE: 11/13/2001

PATENT APPLICATION: US/09/192,579B

TIME: 13:47:12

Input Set : A:\Sequence Listing.txt

Output Set: N:\CRF3\11132001\I192579B.raw

```

287 Lys Ala Glu Gly Tyr Leu Glu Ala Ala Thr Ser Arg Tyr Asn Glu Leu
288   1           5           10           15
290 Val Glu Arg Gly Glu Ala Ala Leu Glu Arg Leu Arg Ser Gln Gln Ser
291           20           25           30
293 Phe Glu Glu Val Ser Ala Pro Ala Glu Gly Tyr Val Asp Gln Ala Val
294           35           40           45
296 Glu
300 <210> SEQ ID NO: 19
301 <211> LENGTH: 1097
302 <212> TYPE: DNA
303 <213> ORGANISM: Artificial Sequence
305 <220> FEATURE:
306 <223> OTHER INFORMATION: Description of Artificial Sequence: DNA sequence
307   of the BCG gene coding for HBHA
309 <220> FEATURE:
310 <221> NAME/KEY: CDS
311 <222> LOCATION: (331)..(924)
312 <223> OTHER INFORMATION: CDS from 811 to 828, from 829 to 846, from 847 to
313   864, from 865 to 885 and from 895 to 915 : peptide
314   which may be particularly involved in interactions
315   with sulphated glycoconjugates
317 <400> SEQUENCE: 19
318 cggctggcgg gtaatcaaac ctgaaggaca gtcattctggg tgaggctcgac cgcaggctga 60
320 tccagccgat cggccggcgc tggccaacag cgactccgtc gatgacgtgc agcaaaggag 120
322 acatgtagtg accggatcag ctgggcctga catctacgaa ctgaccgac aaccgaccgc 180
324 acgatcagga ggtttccccc gcaagtcgcg tgccatgtca atccgcgggt cttgactagt 240
326 cctccctgga ggagccgacg cttgcccaca cgtccagacc aaagatgtaa gaacgccgat 300
328 atcagaaaat agttaatgaa aggaataccc atg gct gaa aac tcg aac att gat 354
329                               Met Ala Glu Asn Ser Asn Ile Asp
330                               1           5
332 gac atc aag gct ccg ttg ctt gcc gcg ctt gga gcg gcc gac ctg gcc 402
333 Asp Ile Lys Ala Pro Leu Leu Ala Ala Leu Gly Ala Ala Asp Leu Ala
334   10           15           20
336 ttg gcc act gtc aac gag ttg atc acg aac ctg cgt gag cgt gcg gag 450
337 Leu Ala Thr Val Asn Glu Leu Ile Thr Asn Leu Arg Glu Arg Ala Glu
338   25           30           35           40
340 gag act cgt acg gac acc cgc agc cgg gtc gag gag agc cgt gct cgc 498
341 Glu Thr Arg Thr Asp Thr Arg Ser Arg Val Glu Glu Ser Arg Ala Arg
342           45           50           55
344 ctg acc aag ctg cag gaa gat ctg ccc gag cag ctc acc gag ctg cgt 546
345 Leu Thr Lys Leu Gln Glu Asp Leu Pro Glu Gln Leu Thr Glu Leu Arg
346           60           65           70
348 gag aag ttc acc gcc gag gag ctg cgt aag gcc gcc gag ggc tac ctc 594
349 Glu Lys Phe Thr Ala Glu Glu Leu Arg Lys Ala Ala Glu Gly Tyr Leu
350           75           80           85
352 gag gcc gcg act agc cgg tac aac gag ctg gtc gag cgc ggt gag gcc 642
353 Glu Ala Ala Thr Ser Arg Tyr Asn Glu Leu Val Glu Arg Gly Glu Ala
354           90           95           100
356 gct cta gag cgg ctg cgc agc cag cag agc ttc gag gaa gtg tcg gcg 690

```

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/192,579B

DATE: 11/13/2001

TIME: 13:47:13

Input Set : A:\Sequence Listing.txt

Output Set: N:\CRF3\11132001\I192579B.raw

L:69 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3

L:87 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:4